351, 290-296). Besides primary anchors, there are also secondary anchor residues occupied in more shallow pockets (Matsumura, M. *et al.*, (1992) *Science* 257, 927-934). In total, six allele-specific pockets termed A-F have been characterized (Saper, M.A. *et al.*, (1991) *J. Mol. Biol.* 219, 277-312; Latron, F. *et al.*, (1992) *Science* 257, 964-967). The constitution of these pockets varies in accordance with the polymorphism of class I molecules, giving rise to both a high degree of specificity (limited cross reactivity) while preserving a broad binding capacity.

In contrast to HLA class I binding sites, class II sites are open at both ends. This allows peptides to extend from the actual region of binding, thereby "hanging out" at both ends (Brown. J. *et al.*, (1993) *Nature* 364, 33-39). Class II HLAs can therefore bind peptide ligands of variable length, ranging from 9 to more than 25 amino acid residues. Similar to HLA class I, the affinity of a class II ligand is determined by a "constant" and a "variable" component. The constant part again results from a network of hydrogen bonds formed between conserved residues in the HLA class II groove and the main-chain of a bound peptide. However, this hydrogen bond pattern is not confined to the N- and C-terminal residues of the peptide but distributed over the whole of the chain. The latter is important because it restricts the conformation of complexed peptides to a strictly linear mode of binding. This is common for all class II allotypes. The second component determining the binding affinity of a peptide is variable due to certain positions of polymorphism within class II binding sites. Different allotypes form different complementary pockets within the groove, thereby accounting for subtype-dependent selection of peptides, or specificity. Importantly, the constraints on the amino acid residues held within class II pockets are in general "softer" than for class I. There is much more cross reactivity of peptides among different HLA class II allotypes. Unlike for class I, it has been impossible to identify highly conserved residue patterns in peptide ligands (so-called motifs) that correlate with the class II allotypes.

The different characteristics of class I and class II MHC molecules are responsible for specific problems associated with the prediction of potential T-cell epitopes. As discussed before, class I molecules bind short peptides that exhibit well-defined residue type patterns. This has led to various prediction methods that are based on experimentally determined statistical preferences for particular residue types at specific positions in the peptide. Although these methods work relatively well, uncertainties associated with non-conserved positions limit their accuracy. Prediction methods for MHC class II-mediated T-cell epitopes essentially follow the same strategy, but are hampered by the fact that the binding groove is open. The latter makes it difficult to locate, in a pool of peptides identified as binders, the 9-residue segment

7

structures may allow the prediction of affinities of peptide for MHC molecules in complexes for which no or only partial MHC molecule structures exist. Since more MHC molecules are known than structures have been experimentally solved, the use of modeled structures allows the prediction of otherwise unobtainable complex affinity data, filling the growing need for such
5   information.

In another embodiment of the present invention the ensemble of step (b) is generated by a computer modeling method, said method being able to generate multiple energetically favorable backbone configurations in relation to the MHC molecule. The use of modeling to generate said ensemble allows the available conformational space to be sampled efficiently,
10   for example in a fashion that is specific for the sequence of said peptide. This provides validation for allowable conformations, and may also provide a more accurate assessment of properties of the complex.

In another embodiment of the present invention the representation of step (b) is retrieved from a library of peptide structures pre-oriented in relation to the MHC molecule. The
15   use of a library provides the opportunity of a drastic reduction of the computational time per peptide since an alternative is to use simulations which may be extremely demanding in computing time due to the large search space. An indirect advantage is the fact that the prediction accuracy can be improved because a large number of pre-oriented peptide structures may be retrieved, and more attention can be paid to the important side-chain
20   placement and affinity prediction steps.

In yet another embodiment of the present invention a complex within said ensemble of step (c) is obtained from a side-chain placement algorithm. The use of a side placement algorithm decouples the side-chain from the main-chain sampling so providing an opportunity to increase the speed and accuracy of the calculation.

25   In yet another embodiment of the present invention the side-chain placement of step (c) not only involves placing the side-chains of the peptide itself, but also involves placing one or more side-chains of said MHC molecule that are in contact with said peptide. The use of both a side-chain placement for peptide and MHC molecules provides the opportunity to generate more accurate models and hence to increase the accuracy of the predicted affinity of the
30   complex.

In yet another embodiment of the present invention a complex within said ensemble of step (c) is obtained from a side-chain placement algorithm suited for global side-chain

8

optimization. The globally optimal placement of side-chains generally yields more accurate predictions compared to local optimization.

In yet another embodiment of the present invention the side-chain placement algorithm of a method above comprises a dead-end elimination (DEE) algorithm, characterized in that said DEE algorithm eliminates rotameric conformations on the basis of a mathematical criterion that allows the detection of conformations that are not compatible with the globally optimal conformation. The DEE approach is helpful in solving the combinatorial search problem by reducing the number of possible rotamers which need to be tested, thereby greatly increasing the speed of global side-chain optimization.

In yet another embodiment of the present invention the side-chain placement algorithm of a method above comprises a FASTER algorithm (Desmet J. et al. (2002) *Proteins* 48, 31-43), said algorithm being characterized essentially by a repeated perturbation, relaxation and evaluation step. The FASTER algorithm improves the side-chain prediction accuracy at a low computational cost, and hence makes provision for more accurate predictions of binding affinity.

In yet another embodiment of the present invention the binding affinity of step (d) of a method above is represented by a single scoring value for the whole ensemble of MHC/peptide complexes, said scoring value comprising the sum of the conformational entropy for the complete ensemble of MHC/peptide complexes, and the average of the said energetical components of each of the complexes of said ensemble. Conformational entropy is a fundamental property of a complex that is preferably computed from an ensemble of structures. The explicit inclusion of conformational entropy contributes in a favorable way to the correlation between predicted and experimental affinities. Furthermore, the incorporation of significant energetic components, in combination with an entropical component, allows a more accurate assessment of the affinity of the complex.

In yet another embodiment of the present invention the binding properties of step (d) of a method above are evaluated for the global complex, thereby accounting for interactions between pairs of residues from the peptide, the MHC molecule and both the peptide and the MHC molecule. The use of global scoring which accounts for interactions between said pairs of residues provides a more accurate assessment of the global energy of the system and hence provides a more exact measure of the affinity of the complex.